CLAIMS:

1. A method for testing validity of a prediction model based on an original data set, comprising:

specifying a test statistic formula;

computing a numerical value NTS of the test statistic using the test statistic formula and the original data set;

specifying a probability distribution relating to the original data set;

creating a plurality of random data sets RDB(i) using randomly generated data, in which i is a positive integer;

computing a plurality of numerical values TS(i) of the test statistic corresponding to the plurality of random data sets RDB(i), and storing each numerical value TS(i) in a numerical test statistic array; and

comparing the numerical value NTS with the numerical test statistic array to determine a non-empty set of percentile values corresponding to the numerical value NTS and an associated non-empty set of percentile indices.

2. The method for testing validity of a prediction model according to claim 1, in which each of the plurality of data sets RDB(i) is distributed according to the probability distribution.

3. The method for testing validity of a prediction model according to claim 2, in which each the data sets RDB(i) has a size that is functionally equivalent to a size of the original data set.

4. The method for testing validity of a prediction model according to claim 1, further comprising:

determining a null hypothesis defining a potential relationship among data in the original data set; and

rejecting the null hypothesis as not accurately representing the original data set when the value of a function of the non-empty set of percentile indices, associated with the non-empty set of percentile values, which correspond to the numerical value NTS, is in an extreme range, indicating that the numerical value NTS did not arise by chance.

5. The method for testing validity of a prediction model according to claim 1, in which the non-empty set of percentile values comprises the greatest percentile value less than NTS and the smallest percentile value greater than NTS, and the non-empty set of percentile indices comprises the two percentile indices corresponding to the two percentile values of the non-empty set of percentile values.

6. The method for testing validity of a prediction model according to claim 1, in which one percentile index is selected, when the corresponding percentile value meets a predetermined criterion for proximity to the numerical value NTS of the test statistic corresponding to the original data set.

7. The method for testing validity of a prediction model according to claim 4, in which the function of percentile indices is a linear combination of the non-empty set of percentile indices.

8. The method for testing validity of a prediction model according to claim 1, in which the test statistic comprises a function of prediction error.

9. The method for testing validity of a prediction model according to claim 4, in which the extreme range comprises one of above a 97.5^{th} percentile and below a 2.5^{th} percentile.

10. The method for testing validity of a prediction model according to claim 1, in which creating the plurality of random data sets RDB(i) comprises using randomly generated data according to a Monte Carlo technique.

11. The method for testing validity of a prediction model according to claim 1, further comprising constructing a confidence interval for the test statistic.

12. The method for testing validity of a prediction model according to claim 1, in which each of the plurality of data sets RDB(i) has the same size, dimension and distribution as the original data set.

13. A computing apparatus for analyzing an original data set, the original data set having a first size, dimension and distribution, the computing apparatus comprising:

a computing device for executing computer readable code;

an input device for receiving data, the input device being in communication with the computing device;

at least one data storage device for storing computer data, the data storage device being in communication with the computing device; and

a programming code reading device that reads computer executable code, the programming code reading device being in communication with the computing device;

the computer executable code causing the computing device to generate a plurality of random data sets, each random data set having a second size, dimension and distribution relating to the original data set; calculate a plurality of numerical values of test statistics corresponding to the plurality of random data sets, each numerical value being calculated according to a test statistic formula; and determine a relationship between the plurality of numerical values and the numerical value of the test statistic

corresponding to the original data set, calculated in accordance with the test statistic formula.

14. The computing apparatus according to claim 13, in which the second size, dimension and distribution is the same as the first size, dimension, and distribution.

15. The computing apparatus according to claim 13, in which the second size of each random data set is functionally equivalent to the first size of the original data set.

16. The computing apparatus according to claim 13, in which the relationship between the plurality of numerical values and the numerical value corresponding to the original data set indicates whether the original data set is characterized by at least one factor that is not based on chance.

17. The computing apparatus according to claim 13, in which determining the relationship between the plurality of numerical values and the numerical value corresponding to the original data set comprises:

determining a plurality of percentile values, based on the plurality of numerical values, and a plurality of percentile indices corresponding to the plurality of percentile values; and

determining a non-empty set of selected percentile indices from the plurality of percentile indices, corresponding to the plurality of random data sets, by determining a non-empty set of percentile values from the plurality of percentile values which meets a predetermined criterion for proximity to the numerical value of the test statistic corresponding to the original data set.

18. The computing apparatus according to claim 17, the computer executable code further causing the computing device to select two percentiles indices,

corresponding to the greatest percentile value less than the numerical value of the test statistic corresponding to the original data set, and the smallest percentile value greater than the numerical value of the test statistic corresponding to the original data set.

19. The computing apparatus according to claim 17, the computer executable code further causing the computing device to select one percentile index when the corresponding percentile value meets a predetermined criterion for proximity to the numerical value of the test statistic corresponding to the original data set.

20. The computing apparatus according to claim 17, the computer executable code further causing the computing device to determine that the numerical value of the test statistic corresponding to the original data set did not arise by chance when the value of a predetermined function of the selected percentile indices is outside a predetermined range of the plurality of percentile indices indicating numerical values that did arise by chance.

21. The computing apparatus according to claim 20, in which the predetermined function of percentile indices is a linear combination of the corresponding percentile indices.

22. The computing apparatus according to claim 13, in which the computer executable code further causes the computing device to construct a confidence interval for the test statistic.

23. The computing apparatus according to claim 13, in which generating the plurality of random data sets further comprises generating the random data sets according to a Monte Carlo technique.

24. A computer readable medium storing a computer program that determines a likelihood of at least one factor in an original data set not arising by chance, in accordance with a predetermined test statistic formula, the original data set having a first size, dimension and distribution, the program comprising:

a calculating source code segment that calculates a plurality of numerical values of test statistics corresponding to a plurality of randomly generated data sets, calculated in accordance with the predetermined test statistic formula, each randomly generated data set having a second size, dimension and distribution relating to the original data set;

a comparing source code segment that compares a numerical value of a test statistic calculated in accordance with the predetermined test statistic formula and calculated with the original data set, with the plurality of numerical values corresponding to the plurality of randomly generated data sets; and

a determining source code segment that determines that at least one factor in the original data set did not arise by chance when the numerical value of the test statistic calculated from the original data set is not within a range, within the plurality of numerical values corresponding to the plurality of randomly generated data sets, representative of numerical values arising by chance.

25. The computer readable according to claim 24, in which the second size, dimension and distribution is the same as the first size, dimension and distribution.

26. The computer readable according to claim 24, the program further comprising:

a percentile determining source code segment that determines a plurality of percentile values, based on the plurality of numerical values, and a plurality of percentile indices, corresponding to the plurality of percentile values;

wherein the comparing source code segment compares the numerical value of the test statistic corresponding to the original data set with the plurality of numerical values corresponding to the plurality of randomly generated data sets by determining a non-empty set of selected percentile indices from the plurality of percentile indices corresponding to the plurality of random data sets associated with a non-empty set of the percentile values from the plurality of percentile values which meets a predetermined criterion for proximity to the numerical value of the test statistic corresponding to the original data set.

27. The computer readable according to claim 26, in which the range of values is based on the plurality of associated percentile indices.

28. The computer readable according to claim 24, in which the second size of each randomly generated data set is functionally equivalent to the first size of the original data set.

29. The computer readable according to claim 27, in which the second size of each randomly generated data set is functionally equivalent to the first size of the original data set.

30. The computer readable according to claim 24, the program further comprising:

a distribution determining source code segment that determines the distribution of the original data set by comparing the original data set with a plurality of theoretical distributions.

31. The computer readable according to claim 24, the program further comprising:

a distribution determining source code segment that determines the distribution of the original data set by sorting the data into bins along at least one dimension.

32. The computer readable according to claim 24, in which the first distribution is not a normal distribution.

33. The computer readable according to claim 24, the program further comprising a confidence interval source code segment that constructs a confidence interval for the test statistic.

34. The computer readable according to claim 24, the program further comprising a distribution determining source code segment that determines an empirical distribution of the original data set.